Seeking coherent explanations — a fusion of structured connectionism, temporal synchrony, and evidential reasoning

Lokendra Shastri and Carter Wendelken
International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, CA 94704

Abstract

A connectionist model capable of performing rapid inferences to establish explanatory and referential coherence is described. The model's ability to perform such inferences arises from (i) its structure, (ii) its use of mutual inhibition among "sibling" types, entities, and rules, (iii) the use of temporal synchrony for representing dynamic bindings, and (iv) its ability to rapidly modify weights in response to convergent activity.

Introduction

Consider the following simple narrative: "John fell in the hallway. Tom had cleaned it. He got hurt." Upon hearing the above narrative most of us would infer that Tom had cleaned the hallway, John fell because he slipped on the wet hallway floor, and John got hurt because of the fall. These inferences allow us to establish causal and referential coherence among the events and entities involved in the narrative. They help us explain John's fall by making plausible inferences that the hallway floor was wet as a result of the cleaning and John fell because he slipped on the wet floor. They help us causally link John's hurt to his fall. They help us determine that "it" in the second sentence refers to the hallway, and "He" in the third sentence refers to John, and not to Tom. Empirical data strongly suggests that inferences required to establish referential and causal coherence occur automatically during language understanding (see e.g., Just & Carpenter 1977; Keenan, Baillet, and Brown 1984; Kintsch 1988; McKoon & Ratcliff 1980, 1992; Potts, Keenan, & Golding, 1988).

Any system that attempts to explain our ability to establish causal coherence during language understanding must possess a number of properties: First, such a system must be representationally adequate. It must be capable of encoding specific facts and events and expressing general regularities (`aka` rules) that capture the causal structure of the environment. In particular, the system should be capable of encoding context-dependent and evidential cause-effect relationships. Second, the system should be inferentially adequate, that is, it should be capable of drawing a range of explanatory inferences by combining evidence and arriving at *coherent* interpretations that are quasi-optimal with reference to a cost-function (Hobbs et. al, 1993). Third, the system should be capable of establishing referential coherence. In particular, it should be able to unify entities and events by recognizing that multiple designations might refer to the same entity or event. Fourth, the system should be capable of learning and fine-tuning its causal model based on experience, instruction, and exploration. Finally, the system should be scalable and computationally effective. The causal model underlying human language understanding would be extremely large. Yet we understand language at the rate of several hundred words per minute (Just & Carpenter 1977). Hence, a system for establishing causal coherence should also be capable of encoding a large causal model and rapidly performing the requisite inferences within fractions of a second.

This paper describes several key extensions to the connectionist model SHRUTI that enable it to draw the sorts of inferences described above. SHRUTI is a neurally plausible system capable of expressing causal knowledge involving n-place relations, limited quantification, and type restrictions. It encodes specific events as well as context-sensitive priors over events. It expresses dynamic bindings via the synchronous firing of appropriate node clusters and performs inferences via the propagation of rhythmic activity over node clusters. This propagation amounts to a parallel breadth first activation of the underlying causal graph, and hence, the reasoning in SHRUTI is extremely fast. The use of weighted links and activation combination functions at nodes allow SHRUTI to encode soft rules and perform evidential inference. SHRUTI supports supervised learning which allows it to fine-tune its causal model in a data-driven manner (Shastri & Ajjanagadde, 1993; Shastri & Grannes, 1996; Shastri, 1999; Shastri & Wendelken, 1999; Wendelken & Shastri, 2000).

In order to carry out inferences for establishing referential and causal coherence, however, SHRUTI's core functionality had to be extended in a number of ways. These include the ability to (i) unify entities and relational instances (events) (ii) posit the existence of entities that are left implicit in the utterance, and (iii) favor interpretations that are more plausible and more likely over others that are less so. These functional extensions were realized in part by introducing mutual-exclusion clusters in the encoding of types and entities and by modifying the behavior of node-types. But more importantly, SHRUTI's inferential behavior was modified by (i) introducing inhibitory interactions among rules sharing a common consequent (effect) and (ii) modeling *short-term-potentiation*, a biological phenomena whereby synaptic strengths (link weights) undergo rapid but short-lived changes in response to convergent activity. Both these changes play a critical role in favoring coherent and more-likely interpretations over less coherent and less likely ones.

The rest of the paper is organized as follows. The next section presents SHRUTI's basic representational machinery. This is followed by an elaboration of evidential reasoning in SHRUTI. Next we discuss mechanisms particularly aimed at the problem of establishing coherence and illustrate the functioning of the model with the help of an example.

SHRUTI's representational machinery

Figure 1 illustrates the encoding of the following fragment of knowledge (expressed in SHRUTI's input syntax):

(1) ∀ x:Agent, y:Location [slip(x,y) ⇒ fall(x,y) (600,900)];
(2) ∀ x:Agent, y:Location [trip(x,y) ⇒ fall(x,y) (800,900)];
(3) *TF: trip(Person, Location) 100;
(4) *TF: slip(Person, Location) 50;

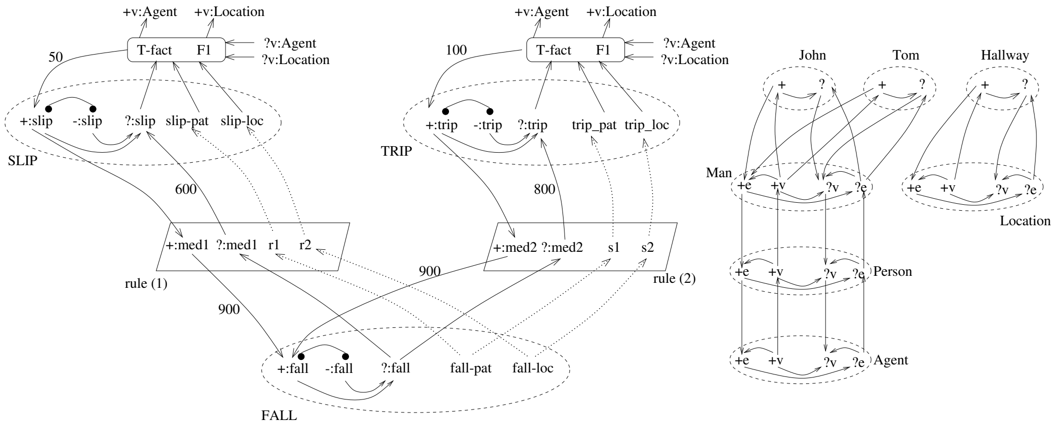

Figure 1: An example SHRUTI network encoding two rules (i) ∀ x:Agent, y:Location [slips(x,y) ⇒ falls(x,y) (600,900)]; and (ii) ∀ x:Agent, y:Location [trips(x,y) ⇒ falls(x,y) (800,900)]; two T-facts, F1 and F2; and a type hierarchy fragment. Links between mediator and type structures, and inhibitory links between sibling rules, entities, and types, have been omitted.

(5) is-a(John, Man); (6) is-a(Tom, Man);
(7) is-a(Man, Person); (8) is-a(Person, Agent);
(2) is-a(Hallway, Location);
Items (1–2) are rules, items (3–4) are taxon-facts (T-facts), and item (5–9) are assertions about types. The first rule states that when an entity of type *Agent* slips at a location, then the latter may fall at that location. The weights (a,b) associated with a rule have an evidential interpretation and we discuss this in the section on evidential reasoning. The weight associated with a T-fact is indicative of the prior probability of the specified event type. All weights lie in the interval [0,1000].

Encoding Relations, Entities, and Types

Each relation is represented by a focal cluster depicted by a dotted ellipse in Figure 1. Consider the focal cluster for *slip*. This cluster includes an enabler node labeled *?:slip*, two collector nodes labeled *+:slip* and *-:slip*, and two role nodes labeled *slip-pat* and *slip-loc* for its two roles *patient* and *location*. In general, the cluster for an *n*-place relation contains *n* role nodes. The positive and negative collectors are mutually inhibitory (inhibitory links are depicted by filled circles).

Assume that the roles of *slip* have been dynamically bound to some fillers and thereby represent an active instance of *slip* (we will see how, shortly). The activation level of *?:slip* indicates the strength with which the system is seeking an explanation for the currently active instance of *slip*. The activation levels of *+:slip* and *-:slip* encode graded beliefs about currently active instance of *slip* ranging continuously from *no* on the one extreme (only *-:slip* is active), to *yes* on the other (only *+:slip* is active), and *don't know* in between (neither collector is very active). If both the collectors receive comparable and strong activation then both collectors can be active, despite mutual inhibition. This signals a contradiction.

The collector nodes of each relation are connected to the enabler node of the relation. For example, *+:fall* and *-:fall* are connected to *?:fall*. These links cause *?:fall* to become active whenever *+:fall* or *-:fall* become active. In effect, these links cause any active assertion about a relation to lead to a query about the assertion. Thus the system continually seeks an explanation for active assertions. The weight on the link from *+:fall* (*-:fall*) to *?:fall* is inversely proportional to the probability of occurrence (non-occurrence) of an instance of *fall* — the less likely an event, the stronger the search for an explanation of the event.

The encoding of types and instances is illustrated at the right of Figure 1. The focal cluster of each entity, *A* consists of a *?:A* and a *+:A* node. In contrast, the focal cluster of each type, *T* consists of a pair of *?* (*?e:T* and *?v:T*) and a pair of + nodes (*+e:T* and *+v:T*). While the nodes *+v:T* and *?v:T* participate in expression of knowledge (facts and attributes) involving the whole type *T*, the nodes *+e:T* and *?e:T* participate in the encoding of knowledge involving particular instances of type *T*. Thus the pair of *v* nodes and the pair of *e* nodes signify universal and existential quantification, respectively. The *levels* of activation of *?:A, ?v:T, and ?e:T* nodes signify the strength with which information about entity *A*, type *T*, and an instance of type *T*, respectively, is being sought. Similarly, the *levels* of activation of *+:A*, *+v:T*, and *+e:T* signify the degree of belief that the entity *A*, the type *T*, and an instance of type *T*, respectively, play appropriate roles in the current situation.

Nodes are computational abstractions and correspond to *small ensembles of cells*, and a connection between nodes corresponds to several connections from cells in one ensemble to cells in the other. Phasic nodes, of which role nodes are an example, produce output spikes in synchrony with their inputs. Temporal-and nodes, such as the enablers and collectors, integrate activity over a broader time window and produce wider output pulses (such a pulse may be identified with recurring high-frequency bursts of spikes).

Dynamic bindings

The *dynamic* encoding of a relational instance corresponds to a *rhythmic* pattern of activity wherein bindings between roles and entities are represented by the *synchronous* firing of appropriate role and entity nodes (von der Malsburg 1981; Shastri & Ajjanagadde 1993; Hummel & Holyoak 1997). With reference to Figure 1, the dynamic representation of the relational instance (*fall*: ⟨*fall-pat*=*John*⟩, ⟨*fall-loc*=*Hallway*⟩) (i.e., "John fell in the Hallway") will involve the synchronous firing of *+:John* and *fall-pat*, and the synchronous firing of *+:Hallway* and *fall-loc*. The entities *+:John* and *+:Hallway* will fire in distinct phases.

Encoding E-facts and T-facts

SHRUTI encodes two types of facts in its long-term memory: episodic facts (E-Facts) and taxon facts (T-facts). These facts provide closure between the enabler node and the collector nodes. While an E-fact corresponds to a specific instance of a relation, a T-fact corresponds to a distillation or statistical summary of various instances of a relation and can be viewed as coding *prior probabilities*. T-facts are conditioned on the type of role-fillers. Typically, T-facts involving salient role-filler combinations such as [*buy(a-Parent, a-Minivan)* $w1$] (i.e., the prior probability that a parent buys a minivan is $w1$) as well as more generic T-facts such as [*buy(a-Person,a-Car)* $w2$] would be learned. The priors for role-filler combinations not explicitly encoded would be inherited from generic T-facts.

Encoding rules

A rule is encoded via a mediator focal cluster (shown as a parallelogram) that mediates the flow of activity between the antecedent and the consequent clusters.[1] The mediator consists of a collector and an enabler node and as many role-instantiation nodes as there are distinct variables in the rule. The enablers of the consequent relations are connected to the enablers of the antecedent relations via the enabler of the mediator. The (+/–) collectors of the antecedent relations are linked to the appropriate (+/–) collectors of the consequent relations via the collector of the mediator. Each of these enabler and collector links for a rule has a weight. The roles of the consequent relations are linked to the roles of the antecedent relations via the corresponding role-instantiation nodes in the mediator. This linking reflects the correspondence between antecedent and consequent roles specified by the rule.

If a role-instantiation node receives activation from the mediator enabler and a consequent role node, it simply propagates the activity onward to connected antecedent role nodes. If the role-instantiation node receives activity *only* from the mediator enabler it sends activity to the node *?e:T*, where T is the type specified in the rule as the role type. This causes node *?e:T* to become active in an unoccupied phase. Node *?e:T* now conveys this activity to the role-instantiation node which in turn propagates this activity to connected antecedent role nodes. This interaction between the mediator and the type hierarchy, in effect, creates activity corresponding to "Does there exist some role filler of the specified type?" This is the mechanism by which new entities are posited and new phases emerge during the course of inference.

Evidential Reasoning

The interpretation of link weights and activation values is intentionally underspecified in the core SHRUTI model. The goal has been to provide a flexible and expressive representational structure which can be fine-tuned according to specific modeling and task requirements. The following describes a specific interpretation of link weights in terms of probabilities that leads to satisfactory explanatory inferences.

A probabilistic interpretation of weights

Refer to the simplified SHRUTI network shown in Figure 2. The weight of the link from the enabler (?) of a relation to its collector (+) equals the (prior) probability of the occurrence of an instance of the relation. This weight corresponds to the weight of a T-fact associated with the relation. The weight of the link from the collector (+) of a relation to the enabler (?) of the relation is *inversely* proportional to the prior probability of the occurrence of an instance of the relation.

[1] The inclusion of a mediator was motivated, in part, by discussions the author had with Jerry Hobbs.

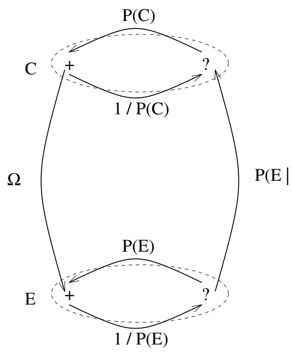

Figure 2: A simplified depiction of SHRUTI's encoding of a rule and T-facts. The rule is $C \rightarrow E$ and the T-facts are the prior probabilities of C and E. The negative collector and all roles nodes have been suppressed.

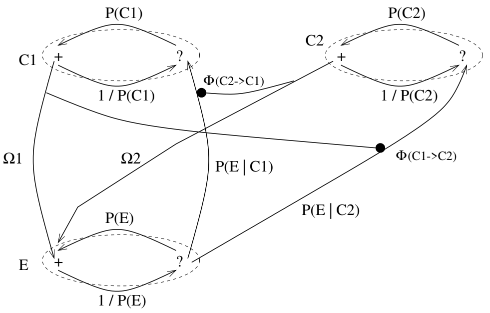

Figure 3: Inhibitory interaction between rules sharing a common consequent.

Now consider the encoding of the rule $C \rightarrow E$. The link weight from *?:E* to *?:C* equal $P(E|C)$, the probability of E given C. The weight, **Ω**, of the link from *+:C* to *+:E* can be interpreted in several ways, as elaborated below. The simplest of these interpretations is $P(E|only\ C)$, the causal strength of C for E (this is essentially the independent component of a noisy-or). Another is $P(C|E)$.

When E is observed to be true, and hence, *+:E*'s activity level is clamped to 1.0, the activation of *?:E* will equal $1 * (1/P(E))$, the activation of *?:C* will equal $(1/P(E)) * P(E|C)$, and that of *+:C* will equal $(1/P(E)) * P(E|C) * P(C)$. A direct application of Bayes Rule shows that the activation of *+:C* reduces to $P(C|E)$ — the desired degree of belief in C under a probabilistic interpretation. If there are multiple causes of E, say $C1$ and $C2$, then subsequent to the clamping of *+:E*, $C1$ and $C2$ will become active at levels $P(C1|E)$ and $P(C2|E)$, respectively, which is again as desired under a probabilistic interpretation (see Figure 3).

Evidence combination

Where there are multiple sources of evidence for some predicate, then we must have some way to combine them. Since each source must communicate independently, along a single weighted link, the approach taken follows that of a belief-net noisy-or (Pearl 1988). However, to allow for more flexible evidence combination within this framework than what a single function can provide, a set of evidence combination functions was developed, based on notions of sufficiency or necessity of factors, and also on degrees of correlation.

Interestingly, these functions suggest several different interpretations of the link weights. At one end of this range is the familiar *noisy-or* function $1 - \prod_i (1 - x_i * w_i)$, where each weight w_i is essentially a measure of the sufficiency of each (independent) potential cause for bringing about the effect. At the other end of the spectrum, a sort of *noisy-and* function $\prod_i (1 - (1 - x_i) * w_i)$ is used where the weight is interpreted as a degree of necessity, the probability that the consequent is false given that the particular antecedent is false (but all other necessary antecedents are true). In between these are *soft-or* (wherein positive correlation is allowed), a set of power averages $((\sum_i X_i^k W_i)/(\sum_i W_i))^{1/k}$ ranging from *max* down to *min* depending on the parameter k, and a *soft-and* analogous to the *soft-or* (see Shastri & Wendelken, 1999).

Mechanisms to support coherence

Several mechanisms have been developed which support the establishment of referential and causal coherence. These include inhibitory connections in the causal model, short-term potentiation, and the ability to create and collapse phases.

Role of inhibitory connections

The encoding of a rule $C \rightarrow E$ in SHRUTI involves inhibitory connections from *+:C* to all the ? to ? links that originate from *?:E* (see Figure 3) and reduce activity at their targets to a degree proportional to the activation of *+:C*. These inhibitory links serve two purposes. First, they provide a mechanism for *contrast enhancement* since they allow stronger explanations to dominate over weaker explanations. Second, they serve the purpose of *explaining away*.[2] It is well known that combining explanatory and predictive inference can lead to problems in an inference system. For example, a system that can infer “John fell” from “John slipped”, and “John tripped” from “John fell” can also have the unfortunate tendency to infer “John tripped” based on “John slipped”. The inhibitory links prevent such unwarranted proliferation of evidence.[3] The precise impact of inhibition depends on the evidence combination function deployed at the site where the inhibitory links converge.

Short-term Potentiation

If *+:fall* receives activity from one of its T-facts it means that *?:fall* is active, and hence, *fall* is being sought as a possible explanation of some event (say, hurt). If at the same time, *+:fall* receives concurrent activity from *+:med1* it means that *fall* is also being predicted as a possible consequence of a *slip* event. In these circumstances, it is *highly likely* that the fall event actually occurred and is both an effect of the slip event and an explanation of the fall event. SHRUTI expresses this increased likelihood via the biologically plausible mechanism of short-term potentiation (STP) (Bliss and Collingridge, 1993). Whenever a collector *+:P* receives activity from one of its T- or E-fact and concurrent activity from a mediator collector node, then the weights of the links from the mediator collector to *+:P* and from the active T-facts to *+:P* increase for a short-duration. Analogous short-term weight changes occur due to convergence of top-down and bottom-up activity at links incident on *-P:* and at *?:P*.

[2]This use of inhibitory connections is motivated in part by Ajjanagadde (1991).

[3]The weights of these inhibitory links can be given a probabilistic interpretation. For example, the weight $\phi(C2 \rightarrow C1)$ in Figure 3 can be viewed as $[P(E) * P(E|C1, C2) * P(C1|C2)]/[P(E|C1) * P(E|C2) * P(C1)]$.

With reference to Figure 3, consider a domain where A is a possible cause of $C1$, and hence we have the rule $A \rightarrow C1$. Now consider a situation where there is independent evidence for A and E and one is interested in determining the probability of $C1$, $P(C1|A, E)$. This probability cannot be exactly computed using only information available locally at node $C1$. Simply combining the evidence arriving from E (i.e., $P(C|E)$) and A (i.e., $P(C1|onlyA)$) using an evidence combination function such as *noisy-or* would typically lead to an underestimation of the correct value. However, the short-term potentiation (STP) of links allows SHRUTI to partially offset this underestimation of the probability of an intermediate relation when both the cause and the effect of a relation are observed. At the same time, the unpotentiated weights continue to propagate the correct probability values when only the cause or only the effect is observed.

At a more global level, STP also has the effect of *priming* the whole subnetwork of nodes and links that constitute a coherent interpretation and creating a strong feedback loop of reverberant activity in a subnetwork of causal knowledge corresponding to a coherent interpretation.

Taken together, the short-term associative increase in weights and the inhibitory interactions leading to the explaining away phenomena, provide a powerful and neurally plausible mechanism that enable SHRUTI to prefer coherent explanations over non-coherent ones.

Mutual exclusion and collapsing of phases

Entities in the type hierarchy can be part of a *phase-level* mutual exclusion cluster (ρ-mex cluster). Consequently, only the most active entity within a ρ-mex cluster can remain active in any given phase. A similar ρ-mex cluster can be formed by mutually exclusive types. Mutual exclusion also occurs in the type hierarchy as a result of inhibitory connections from the + nodes of a type (or an entity) to the ? nodes of all its siblings. This inhibition leads to another sort of "explaining away" phenomenon. If for example, the type query “Is it a Person?” (i.e., activation of *?e:Person*) leads to the queries “Is it a Man?” and “Is it a Woman?”, then strong support received by *+e:Woman* reduces the strength of the query *?e:Man*. In essence, the query “Is it a Man?” is no longer considered important by the system since it was seeking a person and it has already found a woman.

SHRUTI allows separate phases to coalesce into a single phase, or new phases to emerge, as a result of inference. The latter is realized by the allocation of new phases resulting from the interaction between role-instantiation nodes in mediators and the type hierarchy, as described above. The unification of phases is realized in the current implementation by the collapsing of phases based on activity within an entity cluster or within a focal cluster. In the first case, phase collapsing occurs whenever a single entity dominates multiple phases (for example if the same entity comes to be the answer to multiple queries). In the second case, phase collapse occurs if two unifiable instantiations of a relation arise within a focal cluster. For example, the active assertion *+:fall(John,Hallway)* alongside the query $\exists$ *x:Man ?:fall(x,Hallway)* (Did a man fall in the Hallway?) will result in the merging of the two phases for “a man” and “John” via the inferred assertion $\exists$ *x:Man +:fall(x,Hallway)*. The same assertion alongside the query $\exists$ *x:Woman ?:fall(x,Hallway)* would not lead to a similar phase merge because the types Man and Woman are mutually exclusive, and hence, would mutually inhibit one another.

SHRUTI's ability to readily and flexibly instantiate entities and collapse them into a single entity during inference is due to its use of temporal synchrony to represent dynamic bindings.

Simulation Result

The activation trace resulting from the processing of the "John fell" story by a SHRUTI network encoding the rules, T-facts, and type hierarchy described in Section is shown in Figures 4 and 5. Figure 4 shows the actual activation levels of the *+:slip* and *+:trip* nodes as the story is processed by SHRUTI. Figure 5 depicts the activation trace of a larger *subset* of nodes. The depiction in this figure, however, has been simplified to highlight key aspects of the network behavior. In particular, several nodes have been omitted, some intermediate cycles have been omitted and the activation levels of collector and enabler nodes have been discretized to four levels. Please note that due to simplifications made to Figure 5, the time scales along the x-axis in Figures 4 and 5 are not the same. To minimize confusion, we will refer to the times in Figure 4 as cycles and in Figure 5 as steps. The reader may also wish to refer to Figure 1 to ground some of the following description.

Each sentence in the narrative is conveyed to SHRUTI by activating the + node of the appropriate relation and establishing role-entity bindings by the synchronous activation of the appropriate role and entity nodes. The sentences are presented in sequence and after each sentence presentation, the network is allowed to propagate activity for a fixed number of cycles. For example, the first sentence (S1) is communicated to SHRUTI in step 1 (cycle 0) by activating the node *+:fall*, the nodes *fall-pat* and *+:John* in synchrony, and the nodes *fall-loc* and *+:Hallway* in synchrony. The firing of nodes *+John* and *+:Hallway* occupy distinct phases — ρ_1 and ρ_2, respectively.

Activation from the focal cluster for *fall* reaches the mediator structure of rules (1) and (2). Consequently, nodes *r1* and *r2* in the mediator for rule (1) become active in phases ρ_1 and ρ_2, respectively. Similarly, nodes *s1* and *s2* in the mediator of rule (2) become active in phases ρ_1 and ρ_2, respectively. At the same time, the activation from *+:fall* activates *?:fall* which in turn activates the enablers *?:med1* and *?:med2* (the activity of mediator nodes, and role nodes of *slip* and *trip* is not depicted in Figure 5). The activation from nodes *r1* and *r2* reaches the roles *slip-pat* and *slip-loc* in the *slip* focal cluster, respectively. Similarly, activation from nodes *s1* and *s2* reach the roles *trip-pat* and *trip-loc* in the *trip* focal cluster, respectively. In essence, the system has created new bindings for the *slip* and *trip* relations. These bindings together with the activation of the nodes *?:slip* and *?:trip* encode two queries: "Did John slip in the hallway?", and "Did John trip in the hallway?". At the same time, activation travels in the type hierarchy and activates the nodes *?v:Man*, then *?v:Person*, and then *?v:Agent* in phase ρ_1, and the *?v:Location* node in phase ρ_2. The coincident activity of *slip-pat* and *?v:person* node, and the coincident activity of the *slip-loc* and *?v:Location* nodes leads to the firing of the T-fact F1 associated with slip. The activation of F1 causes activation from *?:slip* to flow to *+:slip*. The T-fact F2 associated with trip also becomes active in an analogous manner and conveys activation from *?:trip* to *+:trip*. The level of these activations is a measure of the probability that a person may slip and fall, respectively. At this time, "John tripped" is believed to be a more likely explanation of "John fell" than "John slipped."

While the activation spreads "backwards" from the *fall* focal cluster in the manner described above, activation also travels "forwards" to the *hurt* focal-cluster (not shown in Figure 1) as a result of the encoding of rule (iii) (also not shown) and leads to the weak prediction that John got hurt.

The introduction of sentence S2 in step 6 (Figure 5) (cycle 40 Figure 4) results in the instantiation of *clean* with the bindings *($\langle$clean-agt=+:Tom$\rangle$, and $\langle$clean-loc=+e:Location$\rangle$)*. As a result, Tom gets active in phase ρ_3 and *+e:Location* in phase ρ_4. Note that now we have two instantiations of a location. The second instantiation gets merged with the first (Hallway) as a result of phase merging. This happens in step 8 (see activity of *+e:location* in Figure 5). The pressure for this merging comes from the strong compatibility, and hence, the strong coherence between the activity of hallway and the new location. Note that in the ongoing activity, hallway and the new location (say, Loc1) are active in parallel assertions such as: "John fell on the hallway floor", "The hallway floor might have been wet", "The hallway floor might have been cleaned" and "The Loc1 floor was cleaned" "The Loc1 floor might be wet", "John might have fallen in the hallway floor." At this time, *+:wetFloor* also becomes active as a result of activity arriving from *+:clean* via the mediator of rule (4) (cleaning leads to a wet floor).

By step 10 (Figure 5) *+:slip* becomes more active as a result of the high activation of *+:wetFloor*. The effect of "explaining away" kicks in and causes the activation of *+:trip* to go down by step 12. The strength of *+:slip* increases even further due to (i) the potentiation of links from the mediator for rule (4) (walking on a wet floor may cause slipping), (ii) the potentiation of the link from *?:med1* to *?:slip*, and (iii) the effect of explaining away. The effect of these changes on the activation levels of *+:slip* and *+:trip* may be seen more vividly in the detailed trace shown in Figure 4.[4]

S3 is introduced in step 14 (cycle 80) with the binding *($\langle$ hurt-pat=+e:Man$\rangle$)*. This leads to *+e:Man* becoming active in phase ρ_4 and a second dynamic instantiation of *hurt* (in addition to the earlier instantiation resulting from the inference *hurt(John)*). These two instantiations get merged immediately, and phase ρ_4 gets merged with ρ_1 (John), in step 15 as a result of the phase merging described in the previous Section.

Conclusion

SHRUTI shows how explanatory and referential coherence can arise within a neurally plausible system as a result of spontaneous activity in a network. The network's structure reflects the causal model of the environment and when the nodes in the network are activated to reflect a given state of affairs, the network spontaneously seeks coherent explanations. The time taken to perform an inference is simply proportional to the depth of the causal derivation and is otherwise independent of the size of the causal model. The state of coherence is reflected as reverberatory activity around *closed loops*. The system also makes predictive (forward) inferences, but only those predictions that become part of a coherent explanation gain strength and persist. Coherence arises in SHRUTI as a result of (i) inhibitory interactions among sibling entities, types and rules, (ii) short-term increase in link weights resulting from short-term potentiation, and (iii) the dynamic merging and instantiation of entities.

Acknowledgments

This work was partially funded by NSF grants No. 9720398 and N0. 9970890 and subcontracts from Cognitive Technologies Inc. related to ARI contract DASW01-97-C-0038. Thanks to Jerry Feldman, Jerry Hobbs, Marvin Cohen and Bryan Thompson.

[4] If sentence S2 were delayed, the activity in *slip* would lead to the instantiation of an instance of *clean* with an entity of type *agent* being instantiated as a potential filler of the role *clean-agt*. This entity, however, gets unified with *Tom* upon the introduction of S2.

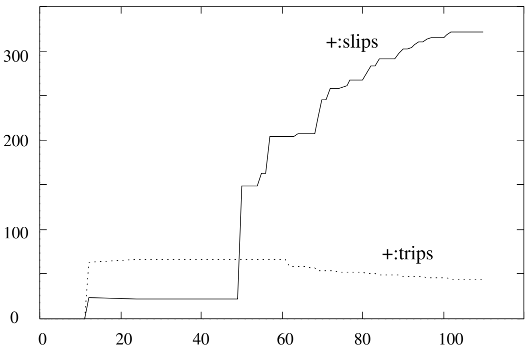

Figure 4: The activation trace of collector nodes *+:slip* and *+:trip* during the processing of the "John fell" story. X-axis is time. The activity of these collectors around cycle 12 is due to associated T-facts. Since tripping is more likely than slipping (100 versus 50), *+trip* has a higher activation. Activity from the *clean* predicate arrives (via *wetFloor*) at the *slip* collector at cycle 50 due to the introduction of S2 at cycle 40, giving *+:slip* a significant boost. >From here onwards the associative weight changes along highly active pathways into *+:slip* result in a large increase in values at around cycle 55. The potentiation of the path from *?:fall* to *?:slip* also contributes to this increase. At the same time, the "explaining away" phenomena leads to the decrease in the activation of *+:trip*. The activity stabilizes around cycle 100. Note that each cycle in SHRUTI roughly corresponds to twice the period of γ band activity, i.e., about 40-50 msecs. (see Shastri & Ajjanagadde 1993).

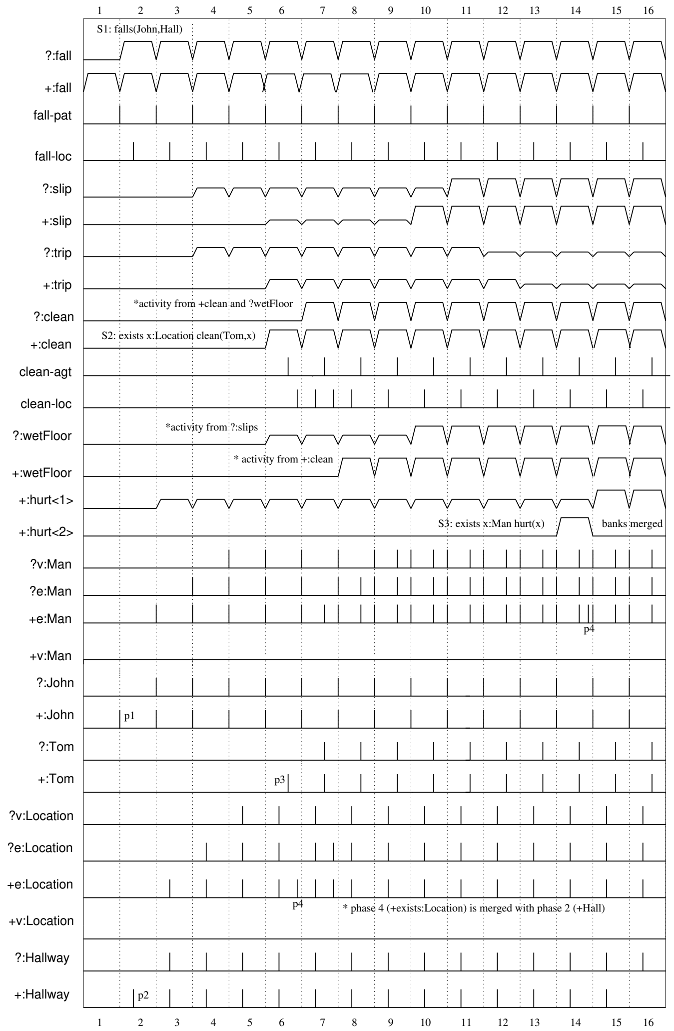

Figure 5: A schematized activation trace of selected nodes.

References

Ajjanagadde, V. (1991) Abductive reasoning in connectionist networks. TR WSI 91-6, Wilhelm-Schickard Institute, University of Tubingen, Tubingen, Germany.

Bliss, T.V.P. and Collingridge, G.L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361, 31–39.

Just, M.A. & Carpenter, P.A. Eds. (1977) *Cognitive processes in comprehension.* Erlbaum.

Hobbs, J., Stickel, M., Appelt, D., & Martin, P. (1993) Interpretation as abduction. *Artificial Intelligence*, 63, 69—142.

Hummel, J. E., & Holyoak, K.J. (1997) Distributed representations of structure: a theory of analogical access and mapping. *Psychological Review*, 104, 427-466.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984) The Effects of Causal Cohesion on Comprehension and Memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115-126.

Kintsch, W. (1988) The Role of Knowledge Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, Vol. 95, 163-182.

McKoon, G., & Ratcliff, R. (1980) The Comprehension Processes and Memory Structures Involved in Anaphoric Reference. *Jrnl. of Verbal Learning and Verbal Behavior*, 19, 668-682.

McKoon, G., & Ratcliff, R. (1992) Inference During Reading. *Psychological Review*, 99, 440-466.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988) Assessing the Occurrence of Elaborative Inferences: Lexical Decision versus Naming. *Journal of Memory and Language*, 27, 399-415.

Shastri, L. (1999) Advances in SHRUTI — A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony, *Applied Intelligence*, 11, 79–108.

Shastri, L. & Ajjanagadde V. (1993) From simple associations to systematic reasoning. *Behavioral and Brain Sciences*, 16:3 p. 417-494.

Shastri, L. & Grannes, D. (1996) A connectionist treatment of negation and inconsistency, *Proc. Eighteenth Conference of the Cognitive Science Society*, San Diego, CA. 1996.

Shastri, L. & Wendelken, C. (1999) Soft Computing in SHRUTI. In *Proc. the Third International Symposium on Soft Computing*, Genova, Italy. June, 1999, pp. 741–747.

von der Malsburg, C. (1981) The correlation theory of brain function. Internal Report 81-2. Department of Neurobiology, Max-Planck Institute for Biophysical Chemistry, Gottingen, Germany.

Wendelken, C & Shastri, L. (2000) Probabilistic Inference and Learning in a Connectionist Causal Network. In *Proc. Neural Computation 2000*, Berlin 2000. To appear.